NATIONAL INSTITUTES OF HEALTH
NATIONAL INSTITUTE OF DIABETES & DIGESTIVE & KIDNEY DISEASES
NATIONAL KIDNEY DISEASE EDUCATION PROGRAM

CREATININE ASSAY AND REPORTING ESTIMATED GFR

Bethesda Suites Marriott
Bethesda, Maryland
January 6, 2003

EXECUTIVE SUMMARY

I. WECOME AND OVERVIEW

Thomas Hostetter, MD, Director of the National Kidney Disease Education Program (NKDEP) at the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) opened the meeting. He welcomed participants and gave them some background on the purpose of the meeting, the magnitude of chronic kidney disease (CKD) and the role of NKDEP. This is the second meeting on this topic. The first meeting held last July was directed at a broader range of GFR estimation and the clinical nephrology and laboratory communities. The present meeting focuses on the creatinine assay with representatives from lab services, instrument makers and those involved in current standardization processes. Several professional societies including the American Society of Pediatric Nephrology, the International Society of Nephrology and the Council of American Kidney Societies have written letters supporting NKDEP's efforts to standardize the measurement of serum creatinine. The National Kidney Foundation also wants to push the standardization effort forward. The goal is to identify those with early kidney disease at the point where something meaningful can be done to treat it. A fact sheet given to meeting participants lists some of the main reasons for focusing on this issue. Some of these reasons are as follows:

_ There are 10-20 million people in the U.S. with chronic kidney disease (CKD) as defined by a GFR less than 60 ml/min/1.73m2 or persistent albuminuria.
_ A simple index of filtration is needed in order to assess those with known CKD; serum creatinine is measured far more frequently in clinical practice than quantitative urinary albumin.
_ There is no single widely accepted creatinine reference method or material. Results for serum creatinine vary widely across clinical laboratories.
_ The most practical way to accurately use estimating equations is to harmonize clinical laboratory results to true creatinine values and to subsequently recast the GFR prediction equations; cystatin may be a better marker but serum creatinine will most likely continue to be widely used for many years prior to the acceptance and use of cystatin. Use of prediction equations will improve recognition of CKD and should therefore improve treatment. The best and most accurate use of the prediction equations depends upon standardization of the creatinine assay.

NKDEP is a new initiative of the NIDDK. This education program is attempting to improve the early diagnosis of kidney disease. End stage renal disease (ESRD) is currently a public health problem. Although economical and effective testing and therapy exist, they are inadequately used. The Federal government covers most of the costs for ESRD treatment. Medicare costs for ESRD are staggering. There were 370,000 cases in 1984; by 2010 there may be 660,000 patients on dialysis. The bill for all payers is 19 billion. Ninety nine thousand people died from ESRD in 2000, second only to lung cancer when compared to all cancer deaths in 2000. The rates for most cancers are stable; some are even decreasing but ESRD rates are rising. Most dialysis patients are dead within 5 years.

Treatment of CKD to delay progression to ESRD is effective. Diabetes fuels most of the increase in ESRD. Intensive glycemic control lessens progression from microalbuminuria in Type I diabetes. However, CKD is not being diagnosed and treated. Only 10% of Medicare patients get annual urine albumin tests, which are recommended to detect early nephropathy. Most patients do not see a nephrologist until 70% of the kidney function has been lost. The diagnosis needs to be made much earlier in the course of the disease. Less than 1/3 of those identified with CKD are put on an ACE inhibitor (ACEi). Tests are misinterpreted if they are done at all, even for those in the hospital. There are several reasons why these patients are missed: kidney disease is asymptomatic in the early stages; the risks are poorly appreciated among primary care physicians who don't realize that they can intervene to slow the progression of the disease; those who are at risk are not tested or the tests are misinterpreted. Most primary care physicians are unaware of prediction equations or they are unable to do the calculations. They also miss patients if they try to estimate the GFR looking at plasma creatinine values. Therapy of kidney disease has improved. The decay in GFR in Type II diabetics was 14 ml/min per year. Those rates of progression can be halved now by treatment to lower than 6 ml/min/year. The time to ESRD depends upon how early the patient starts treatment. The prognosis is better for those who start therapy with a higher GFR. There are also economic reasons to treat CKD. An economic analysis by Trivedi et al in 2002 shows that the earlier kidney disease is detected the greater the savings in costs: if the GFR is less than or equal to 60 the savings is double that of treatment with a GFR of 30 or less.

The goals for this meeting are the following:
_ Demonstrate the value of reporting the estimated GFR with serum creatinine
_ Show the current problems associated with applying the estimating equations
_ Develop the next steps to standardize and improve the accuracy of serum creatinine measurements.

There is a great need to help primary care physicians improve care of those with kidney disease.

II. PRESENTATIONS

K/DOQI and Renal Function Measurement

Andrew Levey, MD, Chief of the Division of Nephrology at Tufts New England Medical Center began this presentation with an outline of the topics to be covered: NKF-K/DOQI Clinical Practice Guidelines in CKD; definition and stages of CKD; measuring GFR; estimating GFR; applications of GFR estimates; reporting estimated GFR; limitations of GFR estimates; limitations of prediction equations; and future tasks. The K/DOQI Guidelines have evolved since the National Kidney Foundation (NKF) launched the Dialysis Outcomes Quality Initiative (DOQI) in 1995. This was followed by the publication of the DOQI Guidelines in 1997. In 2000 the NKF expanded the focus with the Kidney Disease Outcomes Quality Initiative (K/DOQI) and in 2002 they published the CKD Guidelines. These are scientifically rigorous and evidence-based, having been formulated by interdisciplinary work groups. The CKD work group defined and classified the stages of chronic kidney disease, regardless of the underlying cause. CKD is defined as either of the following for ≥3 months: 1) a GFR level <60 ml/min/1.73 m^2, with or without kidney damage; or 2) kidney damage (manifest by kidney biopsy or markers of damage), irrespective of the GFR level. Guidelines 1-3 review the definition, classification, and prevalence estimates for patients with CKD and for individuals at increased risk of developing CKD. Guidelines 4-6 focus on the evaluation of laboratory measurements for the clinical assessment of kidney disease. Guidelines 7-12 describe the association of the level of kidney function with complications of chronic kidney disease, such as hypertension, anemia, malnutrition, bone disease, neuropathy and decreased quality of life. The final guidelines, 13-15, stratify the risk for loss of kidney function and development of cardiovascular disease (CVD). CVD is considered separately because the death rate from cardiovascular disease is higher than that from ESRD in these patients. These guidelines form an evidence-based model for CKD. Three phases of the disease are recognized: CKD, kidney failure and death. A test for microalbuminuria is the best test to find kidney damage from diabetes and hypertension, and is usually abnormal before the GFR is reduced substantially (CKD Stages 1-2). The meeting today is focusing on the stage of decreased GFR (CKD Stages 3-4), which is still before the stage of kidney failure (CKD Stage 5, approximately equivalent to ESRD). There are approximately 8 million adults in the US with CKD Stages 3-4. This large number presents a work load problem for the primary care physicians, cardiologists, geriatric physicians, for every specialty in medicine. The clinical action plan in Guideline 2 shows the stage of kidney disease, corresponding GFR level, and the action plan for that stage. Thus, the K/DOQI CKD guidelines enable the physician to locate the patient on the action plan, to implement a treatment plan to improve outcomes.

The NKF-K/DOQI CKD Guideline 4 says that the level of GFR should be estimated from prediction equations. The wording stating that "the serum creatinine alone should not be used to assess the level of kidney function" attempts to change current clinical practice. It is not enough to review the serum creatinine; GFR function must be estimated from the creatinine measurement. The guideline further states, "autoanalyzer manufacturers and clinical laboratories should calibrate serum creatinine assays according to an international

standard." The 24-hour timed urine collections do not improve the GFR estimate and are not necessary. GFR can be thought of the product of the number of (N) times the single nephron GFR (SNGFR). There are advantages in using GFR as a measure of kidney function. The GFR level is affected by a broad range of factors, it correlates with the severity of the complications of CKD, and it correlates with the degree of glomerular and tubulo-interstitial damage. However, the following are some disadvantages in using GFR: The level may not change despite kidney damage due to an adaptive increase in SNGFR; infants have lower levels and there is an age-related decline in adults, factors other than kidney disease such as diet and drugs, especially antihypertensive drugs may also affect GFR. The gold standard for GFR ascertainment is renal clearance of the ideal filtration marker: inulin. Alternative filtration markers include exogenous ones like iothalamate and iohexol and endogenous ones like creatinine. Assuming that urinary excretion of an endogenous filtration marker (UxV) is constant among individuals and over time within an individual, then the reciprocal of the steady state serum level (Px) is related to GFR. Creatinine, however, is difficult to measure and is not an ideal filtration marker. Prediction equations have been developed to estimate GFR without a urine collection to measure clearance. In analyzing the different equations used to estimate or predict GFR, it appears that the Modification of Diet in Renal Disease (MDRD) study's abbreviated 4 variable equation gives the best estimate. This equation has the additional advantage that it is not necessary to measure the patient's height or weight. If comparing to other equations, it is important to show the result with and without bias. This equation appears to be the best one to use for the present time. It can be improved but it is the best for now.

Dr. Levey presented information on comparing estimates of GFR against measured GFR among the MDRD study baseline cohort. Creatinine clearance tends to overestimate GFR by a small amount. Even when correcting for bias, the 24-hour creatinine clearance was not as accurate as the prediction equations in estimating GFR. This finding is the justification for the statement in Guideline 4 that such collections are not necessary in routine clinical practice. In using prediction equations, it is critical to correct for differences in calibration between the laboratory in which serum creatinine is measured and the laboratory in which the prediction equation was defined. This is especially important when serum creatinine is near the normal range.

An analysis of the prevalence of abnormalities by level of GFR using data from the National Health and Nutrition Survey of 1988-1994 (NHANES III) shows a systematic difference (bias) between the NHANES and MDRD Study laboratories of approximately 0.2 mg/dl. Failure to correct for this difference could lead to GFR estimates that were in error by greater than 20% in individuals with normal serum creatinine. After bias correction, the estimated GFR in NHANES as a function of age was similar to measured GFR in the classic study of Shock and Davies using inulin clearance. Estimated GFR was slightly lower in NHANES, due to the inclusion of women and individuals with hypertension, diabetes and kidney disease, who were excluded in the study by Shock and Davies.

Now, let’s discuss some of the applications of GFR estimation in CKD. An analysis of the level of estimated GFR at the onset of kidney failure was performed by computing

estimated GFR using data from the US Renal Data System (USRDS). Dialysis was initiated at a GFR level <15 ml/min/1.73 m^2 in 98% of patients; approximately 65% of patients start at a GFR less than 10 ml/min/1.73 m^2, and about 25% began dialysis at a GFR of less than 5 ml/min/1.73 m^2. It will be easier to find those needing treatment if GFR is reported rather than simply reporting serum creatinine levels. It is also possible to estimate the years until kidney failure based on the GFR and the rate of GFR decline. As you can see, for patients with GFR >60 ml/min/1.73 m^2, a decline in GFR $\geq$4 ml/min/1.73 m^2 per year results in the onset of kidney failure, defined as a GFR <15 ml/min/1.73 m^2 within 10 years, and can be defined as a "fast" rate of decline. In contrast, the aging-associated decline in GFR is approximately 1 ml/min/1.73 m^2 per year. Based on these facts, we can operationally define the goal of therapies for CKD is to slow the rate of decline of GFR to 2 or 1 ml/min/1.73m^2 per year. We have stated that CKD is associated with an increased risk of CVD. An important research question is whether the increased risk of CVD in individuals with a decreased GFR is independent of other known risk factors for CVD. Analysis of data from the Atherosclerosis Risk in Communities (ARIC) Study shows that individuals with a GFR of <60ml/min/1.73 m2 were found to be at the increased risk of CVD compared to those with a higher GFR. For every 10 ml/min/1.73 m^2 lower GFR, the CVD risk was 5% higher. The data also suggested that GFR may be an independent risk factor. For a GFR within the range of 15-59 ml/min/1.73 m^2, the adjusted relative risk for cardiovascular disease was 1.38. The unadjusted rate was much higher. In practice, the unadjusted relative risk may be more helpful to clinicians; for example, the unadjusted relative risk for heart disease is reported when high serum cholesterol is found.

The New England Medical Center's Lifespan Laboratory has instituted a new requisition slip for a clinical test "GFR Estimate." This terminology to "Estimated GFR," because clinicians need to be able to find the test result under "G" rather than under "E"! The form is simple and has a check box for African Americans as the estimating equation from MDRD uses this racial factor. It may prove more helpful to simply print the result for African Americans and non-African Americans and not worry about collecting information on race. The result does not list a normal range since this differs by age. There is an interpretive comment on the lab result, which is too little information for nephrologists but is a good starting point for general practice. As the methods are refined, the test can still be labeled "GFR Estimate."

In summary, there are a large number of uses for GFR estimates in research and in clinical practice. Regarding research, estimates using the NHANES III data show that GFR declines with age starting at age 20, rather than at age 40 as previously suggested. The Davies and Shock data from the landmark 1950 study with inulin clearances show that men's values are higher than women's. Differences by race have not been systematically studied. The burden of a GFR <60 ml/min/1.73 m^2 is not yet fully known. Some of these research questions can be answered only by measuring GFR. For others, some answers can be provided by estimating GFR. GFR estimates can be used in clinical practice to detect and intervene early in kidney disease, to predict the severity of kidney disease (stages of CKD, association with complications, timing of kidney replacement therapy and drug dosing) and they can predict the interval until kidney failure. There are

limitations in using GFR estimates from serum levels of endogenous filtration markers and prediction equations that include: needing steady state conditions; factors affecting generation, renal handling, extra-renal elimination of the filtration markers; inaccuracy of lab measurement of the filtration marker and problems of generalizability of the prediction equations. In clinical practice the limitations of GFR estimates include the fact that there is physiologic variability with a wide range of normal; there are age-related differences; and the unknown sensitivity and specificity of the GFR cut-off of < 60ml/min/1.73 m2 for detecting kidney disease.

Dr. Levey suggests the following tasks relating to clinical labs need to be done: improve accuracy of serum creatinine measurements; use prediction equations to estimate GFR; report estimated GFR and thereby improve outcomes for patients with CKD. He closes with the following suggestions for research: evaluate new filtration markers and prediction equations to estimate GFR; evaluate alternative definitions for CKD and evaluate burden of CKD using alternative definitions.

Derivation of Estimating EQs

Tom Greene, PhD, of the Department of Biostatistics and Epidemiology at the Cleveland Clinic Foundation spoke about estimating GFR from serum creatinine and developing a multivariable prediction equation in the MDRD study. There is a reciprocal relationship between serum creatinine and GFR: at a given serum creatinine there is a wide variation in GFRs associated with a fixed creatinine so it was deemed useful to develop a new equation. Then it was necessary to determine what predictor variables to use in the equation and to know what mathematical functions to use to predict GFR based on those variables. A consideration in the development of the prediction equation was using a log-transformed model, which was linear on the log scale and equivalent to a multiplicative model and stabilizes the variability at a higher GFR. The equation had to eliminate positive skewness and include some other models as special cases. Most of the other equations developed in the past have this same kind of structure. A wide range of possible variables was reviewed. The best subset selection in the MDRD sample included serum and demographic variables, number variables and predictor variables such as plasma creatinine. The plasma creatinine variable alone accounts for about 80% of the variation. The next important factor is gender, which reduces the error by about one quarter. Race is the next important factor and this reduces error to about .866. Age reduces error to about .875. As variables are added error is reduced a bit more. They compromised on 6 variables in the MDRD study. The simplified MDRD equation has 4 variables: plasma creatinine, gender, race and age. The model then had to be tested. If the same data used to develop the model is used to test it, the performance is overestimated. They therefore separated the data in a training sample (2/3 of the dataset) and a validation sample (1/3 of the dataset). They then compared the predicted vs. the estimated GFR using the validation sample. The diagnostics for the prediction model are as follows: the median 1% error shows greater accuracy of the MDRD equation in generalizing for different renal diseases and demographic characteristics. One of the complications is the fact that iothalamate GFR procedure itself varies greatly in the clearance obtained. Thus, in reviewing different equations for accuracy the results may be overly negative or

pessimistic. That is, some of the deviation between equations and GFR is due to measurement error in GFR. The variance component analysis of repeat GFR measurements at baseline months 0 (beginning of baseline period) and 3 (end of baseline period) suggests that 4.6% of the residual variance is due to measurement error in GFR. Some of the changes in GFR may be true changes in renal function so some of that 4.6% is not random longitudinal variation but real changes in patients who are progressing. Therefore, comparing methods after controlling for GFR measurement error reveals an optimistic assessment of about 5% for the MDRD equations. These equations are accurate only within the MDRD study so a major methodological issue is the extent to which they are generalizable beyond that study. There is no bias due to using the same patients in getting and validating the sample (since they did not do that). If selection bias was present in enrolling the patients, that won't be seen in the equations. Most of the patients were non-diabetic, not high GFRs, etc. So the question about whether or not this 4 variable equation will generalize throughout general medical practice remains open. Estimating GFR from creatinine and other factors depends upon creatinine secretion as well as extrarenal secretion of creatinine and these variables are not usually measured. The MDRD equation is based on the relationship of observed and unobserved characteristics to arrive at a true GFR. They also did external validation of the MDRD equations in the African American Study of Kidney Disease and Hypertension (AASK). Patients in this study had higher GFRs in the range of 70-80 with some in the 90s. One thousand seven hundred three African Americans had hypertensive nephrosclerosis. The laboratory used for measuring serum creatinine and the same methodology for measuring it were used in both AASK and MDRD (mean GFR of 75; mean age of about 50+). The goodness of fit for the MDRD equations and the best equation in AASK was not meaningfully different. The MDRD equation predicted GFR as well as it could be predicted in the context of AASK. These patients like those in MDRD are non-diabetic with varying levels of renal insufficiency. The MDRD equations will have more validation in the next 1-2 years with application in yet more studies measuring true GFR.

Calibrations and the Use of the EQs

Josef Coresh, MD, PhD of the Welch Center for Prevention, Epidemiology & Clinical Research at Johns Hopkins University focuses on prevention in his role as an epidemiologist. He feels it is appropriate to find kidney disease before it progresses to end stage. His presentation reviews estimating GFR, calibration of serum creatinine from a practical perspective and the effect of differences across and between laboratories in the U.S. Dr. Coresh agrees that it is important to establish a standard for creatinine measurement. The cut-off of 60 for GFR is easy to communicate to doctors and to patients. Patients with this GFR level will lose half their normal kidney function. Patients at two standard deviations beyond normal of 60 are around 80 or 90. The difference between 60 and 80 is huge so calibration becomes important. The precision of the GFR estimating equations is better at the lower rather than the higher GFRs. The MDRD study prediction equations yield good results whereas the 24-hour urine collections do not. If the equations are corrected for bias their performance is improved making them about

80% accurate. The Cockroft-Gault equation is easier to memorize but it cannot be traced back to the laboratory and samples where it was developed. The College of American Pathologists (CAP) changed the calibration of serum creatinine late in the NHANES III study. A comparison of frozen samples 8 years later between the levels found at the lab in NHANES III and at the Cleveland Clinic found little bias. Using redundant calibration they estimated that bias between the MDRD lab measurements and the NHANES III lab measurements was .23 mg/dl in a 40-year old African American man. When estimating, it makes a difference whether the creatinine is true or not. The true value becomes more important when creatinine values reach 1.6. In comparing laboratories, differences of .2, .3 or .4 are not uncommon and need to be improved. Estimates of GFR based on low creatinine levels make dietary and body composition factors more important when dealing with moderate elevation in serum creatinine related to moderate changes in GFR. There is a need for an education campaign to teach doctors, patients, laboratories and manufacturers how to test, interpret tests and design treatment based on the test results. There are differences, too, among men, women and children in the normal values. There will be a hurdle to jump to get all on the same page, but it is important to get physicians to act to get early intervention in kidney disease.

There are about 8 million people with a GFR of 60 or less, or 4% of the total population but about 20% of those over age 65. The GFR declines with age but it is not known whether that is acceptable or not. It used to be thought normal to have older people have higher blood pressure numbers. However, it is now known that older people benefit most from lowering blood pressure and suffer fewer strokes because of the interventions. Comorbid conditions like anemia are more prevalent with a GFR over 60; hemoglobin levels in women with a GFR below 60 drop dramatically. A number of abnormalities occur at a GFR less than 60. In the current state of affairs it is not sufficient to know what a reference value of 1.0 means in a particular laboratory. This hinders research as well as practice. Again, calibration is important since not calibrating changes GFR significantly (use serum creatinine minus 0.23). Without calibration a GFR of 90 can come out as a GFR of around 70 in non-diabetics.

In evaluating prediction equations, the AASK and MDRD equations are similar. The line of identity and the regression line are similar. However, Cockroft-Gault uses ml/min, which gives a higher value for a bigger person. It is a higher estimate than the MDRD. This difference is more noticeable at a higher age. The 140 minus age term of the Cockroft-Gault equation is a steeper estimate than in the MDRD. It is important, therefore, to further generalize the MDRD equation and see how it applies to diabetics, obese, etc. in the general population. The equation may change quite a bit in 5 years, but the changes will not cause a paradigm shift. Having laboratory reports that show the estimated GFR will also help family doctors see improvements. The prediction equations are far superior to using serum creatinine alone. New measures should be compared to the state of the art equation using state of the art methodology. More data is needed on the precision of the GFR. It is best to correct for calibration at the source. Information systems are improving and can do the simple calculations for physicians. Furthermore, reasons beyond prediction and progression of kidney disease to report estimated GFR include the need for a good GFR measure to dose medications correctly. It is also critical

to pediatricians and to transplant physicians to have reliable serum creatinine values from laboratory to laboratory around the country. Pediatricians support the Schwartz equation for young patients over the MDRD.

State of the Art and Path Forward in Creatinine Measurement and Standardization

Greg Miller, Ph.D., Professor of Pathology at Virginia Commonwealth University, reviewed the available information on how well laboratories perform creatinine estimates. He finds 3 sources of error: systematic bias (or trueness vs. a reference measurement procedure); random bias due to calibration variability (ability of manufacturers to achieve uniform calibration among individual laboratories using the same method); and random bias from replication or the imprecision of the assay (usually a standard deviation). Method trueness vs. a reference method procedure (RMP) was compared in a survey done by the College of American Pathologists (CAP) in 1994. The National Institute for Standards and Technology (NIST) measured a value of 0.86 mg/dL with an observed bias for six representative method groups ranging from 0.22 to 0.04 mg/dL. The bias in measuring a creatinine level impacts the GFR. Within a method group laboratories are able to get plus or minus 0.2 mg/dL in their ability to put calibration into practice. As the bias of different methods is more positive the GFR gets smaller and the range of variability gets smaller. The systematic bias element needs to be 0 so the same equations can be used with different methods. Data from a multi-laboratory QC program shows a comparison among four different methods. For three methods at a level of 0.7 mg/dL for creatinine there was a range of approximately 0.12 mg/dL between the monthly mean values for QC specimens. One method had a more variable range of monthly mean values in the field. Laboratories could get a mean value which varies from 0.6 to 1.1 mg/dL over a period of several months, which is a substantial change in creatinine due to changes in calibration consistency. The monthly variation in calibration bias was random with some months notably worse than others among laboratories. This is a random bias component of calibration consistency, which differs from a systematic bias. Laboratories recalibrate a various times based on the methodology. Some recalibrate daily, some weekly, some monthly according to the type of method.

Replication imprecision is the next source of error. The monthly standard deviation among the laboratories in the QC program is calculated monthly over 12 months. The within-calibration replication errors are low: ranging from 0 to 0.17 at creatinine of 0.7 and 1.9mg/dL but they still need to be considered. The total error is the sum of systematic bias% plus the square root of the sum of random bias as CV% from calibration variability and replication CV%. Total error in this context ignores contributions from method non-specificity. The Jaffe reaction itself also has limitations. The trueness bias cannot be made equal to 0 through a standardization program when non-specific serum components contribute to the measurement. Non-specificity has to be added on top of the impact of method variability and will be more significant in patients such as diabetics. It is necessary in going forward to define the total error goal necessary for clinical management; make the trueness bias equal to 0; reduce the calibration variability and reduce measurement imprecision to achieve the desired total error. Intermethod standardization must use a commutable material defined as follows: "The same quantity

of measurand in a patient's serum and in a reference material (RM, PT, EQA, AC) have the same numeric result from a particular method." This is not easy to achieve. There are few commutable reference materials on the market. The majority of Proficiency and quality control programs use materials that are not commutable. The following steps are necessary in order to make the trueness bias equal to 0 in the short and long term:

Trueness Bias 0--Short Term	Trueness Bias 0--Long Term
1) Determine bias of routine methods vs. RMP	1) Develop commutable reference material (RM) at 2-3 levels
2) Determine correction to current MDRD equations for each method	2) Establish reference measurement procedure laboratories for value assignment of the RM
3) Revalidate method bias and correction factors annually	3) Method manufacturers use the RM for calibration traceability
4) CAP FFS specimen in October 2003 Survey will measure method bias	4) PT surveillance with FFS to monitor performance of routine methods

The College of American Pathologists (CAP) will include a fresh frozen sample (FFS) of an off-the-clot serum commutable specimen in October of this year (2003) to assess bias among methods and between labs at a level of 1 mg/dL for serum creatinine. Standardization models such as the ones for cholesterol and HbA1c can be used to develop a long-term program for trueness standardization. In the cholesterol program manufacturers split patient specimens between field methods and a reference method to establish proper calibration values for their method specific calibration materials. The HbA1c program introduced proficiency testing with commutable whole blood materials to see how well the standardization program is working. The NCCLS created the C-37 protocol to define a non-supplemented off the clot frozen serum reference materials. This material is commutable among 26 methods for cholesterol and data exists for creatinine but has not been analyzed. It is possible to supplement a similar material with pure creatinine to get higher creatinine levels; pure organic creatinine is unlikely to change the commutability properties of the material. Achieving calibration traceability to the correct creatinine values will produce a lower creatinine and calculating creatinine clearance from serum and urine values will yield a higher clearance number. The reference range for creatinine will also change. Thus education will be necessary to implement a calibration standardization program. The Jaffe method is non-specific and may be advantageous to transition to more specific enzyme-based or other methods. The creatinine reference method is well established as isotope dilution mass spectroscopy (IDMS). It may be necessary to use HPLC as an interim reference method when it has been validated to have a known and constant bias to IDMS.

Two other sources of variability are the method calibrator value assignment specifications and the instrument measurement specifications. To reduce the measurement imprecision methods could do several things: report creatinine with two decimals; improve the analytical imprecision at low values, possibly at the expense of measurement range; and implement more rigorous instrument maintenance protocols.

The final recommendations as this effort goes forward are:
_ Establish a NKDEP laboratory standardization panel

_ Develop specifications for total error for serum/plasma creatinine measurement
_ Develop an infrastructure and technical protocols to achieve inter-method and inter-lab standardization (this will enable: uniform calculated GFR results equivalent to the values achievable with a reference measurement procedure (RMP); and reduced imprecision).

There was some discussion following Dr. Miller's presentation. The CAP fresh frozen serum (FFS) specimen to be sent in October 2003 is going to be one specimen that represents off-the-clot serum from normal individuals with an expected creatinine of 0.9 to 1.0 mg/dL to measure variability among individual laboratories and trueness bias among method groups. It would be ideal to send 2 different specimens but there are insufficient funds at present to cover that expense.

With regard to preparation of a reference material the issue of ownership was raised. It may be possible that NIST could make such a commutable reference material available with a normal and a higher spiked level of creatinine. The specifications for preparation of an unadulterated material already exist in NCCLS C37-A. One of the tasks of the working group or panel would be to validate another pool at a second creatinine level. The C-37 specifications are documented only with regard to cholesterol and that material has two different levels for cholesterol only. There are several HPLC methods in use for creatinine. It is undetermined at present which laboratories could provide reference measurements services. It may be that NIST could provide the material and it may not be necessary to have a group of reference procedure laboratories. Funding is required, however, for NIST to go forward. There are also activities internationally on the ISO level. In Europe there is a directive for diagnostic products requiring manufacturers to trace calibration where possible to a higher order reference procedures or reference material. The effort with NIST on the development of reference materials and methods should take this directive into account. The current MDRD equation can only be used after a method has been correlated to the method in the clinical studies that developed the equation. Calibration may be adjusted to eliminate systematic error, but there may still be issues with methodology, such as different bias at different creatinine levels due to non-specificity, that impact the reliability of the MDRD equation. Other factors such as ethnicity and ages not included in current clinical trials may impact usefulness. A current NIH cohort study of 3000 chronic renal insufficiency patients may help improve these equations in the next 3-4 years or show how robust the equations already are.

III. Reports From Pilot Programs Reporting Estimated GFR

Veterans Affairs-Washington, D.C.

T. G. Patel, MD, MACP, Program Chief for Kidney Diseases, Diabetes and Cancer and Chief Nephrologist with the Department of Veterans Affairs in Washington, D.C. outlined where the VA stands in its effort to implement reporting of estimated GFR (eGFR). The VA is one of the largest healthcare systems with 62 dialysis units across the country and a 22 billion dollar budget. They have a computerized patient record system that is one of the largest in the world. They have been using the Cockroft-Gault equation for reporting eGFR and have their own guidelines on how to manage kidney disease.

They are now switching to the MDRD 4 variable equation and will incorporate the K/DOQI Guidelines into their program. The equation will be explained in the note from the laboratory. The Navy Hospital in Bethesda, Maryland uses the 5 variable MDRD equation. Walter Reed Army Hospital is also planning to implement the 5 variable equation. The VA has the ability to plot graphs with a patient's laboratory values. When they switch to the MDRD equation in a couple of weeks they can compare that equation's eGFR with the Cockroft-Gault. This is on the VA CPRS system. They can also graph laboratory values against time and this becomes part of the patient's medical record. They have the ability to create templates for specialty clinics. There is a national databank for all VA patients in Austin, Texas from which they can extract data. Since graphing is possible, they can do a graph to show a patient how his/her creatinine is rising as the kidney function decreases. Showing a patient an eGFR and explaining how the kidney function is decreasing and the time to dialysis is nearing has a profound effect on the patient. This can also be used for research. The ability to graph reports impacts care and patient understanding. Patients understand a graph of eGFR much better than previous explanations since it is difficult to understand that a rising creatinine means a worsening of the kidney condition. The eGFR number makes it easier to explain when a fistula will be required and the stages of kidney disease can be more easily detailed. The VA has not yet evaluated the response of primary care physicians to receiving an eGFR on laboratory printouts. The general understanding, however, is that they are happy to receive it and there have been no complaints. This will be implemented nationally in a couple of weeks.

In terms of supporting a standard, the bottom line for the VA is money: what will it take to implement it and change current systems? Since they are a single system it is easier to change. All vendors need to join to support a single standard. Data on comparing laboratory values and laboratory parameters among national VA laboratories is not available at the present time and analyzing it will take money. However, Dr. Patel can question the VA laboratories on what they do and how they do it. It is also possible to get CAP data and look at that for all VA hospitals and compare variables among hospitals. Dr. Patel can look at the laboratories that show variability in creatinine and see if the percent of values in certain categories changes as a function of creatinine standardization differences among laboratories. The VA does not routinely capture information on race. Serum creatinine will be reported for African Americans and non-African Americans.

Laboratory Corporation of America

Dr. James Fleming, Associate Vice President and Director of Scientific Affairs for Laboratory Corporation of America spoke about his company's initiatives with the National Kidney Foundation and Coventry Health Care, Inc. in western Pennsylvania. Dr. Bernard Mansheim, Senior Vice President and Chief Medical Officer of Coventry Health Care, Inc. initiated conversations with Laboratory Corporation of America about adopting the K/DOQI Guidelines. These guidelines fit nicely with initiatives of the American Diabetes Association (ADA) and the National Cholesterol Education Program (NCEP). Changes in laboratory systems and practices have a wide impact on such things as client education materials, sales force education, utilization reports, test master, etc. Laboratory

Corporation selected the 4 variable MDRD equation to implement. It requires two different calculations--one for men, one for women--and a race variable. The equation has to be hard coded into the laboratory information system (LIS system). Not all LIS systems are capable of exponential programming. If there is no age or gender listed on the test request form or the patient is less than 18 years, a GFR cannot be calculated. Since information on race is generally not captured they also need to account for the racial differences in GFR. In order to report serum creatinine the company has to calibrate the serum creatinine and correlate their method to the Scr (serum creatinine) method used to develop the MDRD equations. They sent samples from Laboratory Corporation's Columbus laboratory which is the site for the pilot program to the Cleveland Clinic. Laboratory Corporation uses the Roche Rate Jaffe/Modular while the Cleveland Clinic uses the Beckman Rate Jaffe/CX3 Synchron so there is a slight proportional bias. There is always an analytical bias due to the use of different reagents, optics, etc. If one has a result of 1.3 calibrated to the Cleveland Clinic that is almost 1.44 or almost 10 ml/min difference in the calculated GFR. Therefore, calibration is very important. The client report will conform to the K/DOQI Guidelines in terms of the information given. There is a limit for the company of 59 characters per line and 20 lines per page (the EDI interfaces to 3rd party LIS systems are a limiting factor). The laboratory report will have the creatinine result and a calculated GFR. A table will also be on the report explaining the results with and without kidney damage. There is a referral to *kdoqi.org* for further information. The report is still in the process of being modified by the working group. The educational material for this change is very important. Laboratory Corporation is working with the NKF on the development of the educational materials. The materials will go to physicians to help them understand how to read the reports. The Sales Force will also need to be educated. The marketing department uses instructional modules, audiotapes, sales literature, etc.

This pilot program will begin in western Pennsylvania with targeted clients who will be given the technical review and can order it. This is a program with NKF, Dr. Andrew Levey's staff, Laboratory Corporation and Coventry Health Care, Inc. Laboratory Corporation will give utilization reports to Coventry and send it to NKF for assessment. This is basically an administrative before and after review of Coventry at the western Pennsylvania site, which gets patient laboratory results from Laboratory Corporation's Columbus laboratory. At the western Pennsylvania site Laboratory Corporation will report an estimated GFR whether it is ordered or not whenever a creatinine is ordered in those outpatient practices. At another Coventry site it would be worthwhile to compare the before and after trend without implementing the GFR; doctors are allowed to order GFR at these sites. At a third site they will study the effect of implementing GFR reporting plus implementing the full-scale NKF K/DOQI Guidelines to tell providers what to do with the GFR. There are 3 different sites and this will allow a comparison of cross-sectional data at two different times: before and after guideline implementation. This may show what people are doing with the data and how practice patterns change and how patient outcomes change. They will then review the results before launching nationwide. The program in western Pennsylvania will begin in the next 1-2 months. This is in place of a prospectice cohort study that would require more resources.

A question was asked about how heterogeneous the testing platforms are. The answer is that they are all Roche modular but not all Roche modulars work the same so they compare to the test laboratory of their system. If a physician asks for a GFR but does not supply all the needed information the calculation has been programmed to not give an answer. The client can call back and supply the missing data and the GFR will be estimated. Approximately 23 laboratories in the system use the Roche modular system. The variability among them is 5% at .9 and the inter-laboratory variability is between 4-5% monthly. They did this comparison with the Cleveland Clinic at one time with 200 samples about 6 weeks ago. It would be good if such a program could be instituted annually. The GFR is done without charge. There is no financial incentive for the company to calculate it. They are only reimbursed for creatinine.

Kaiser Permanente--Southern California

Mark Rutkowski, MD, a nephrologist with the Southern California Permanente Medical Group spoke about that organization's experience implementing reporting of GFR. Kaiser has been an innovator in the management of early kidney disease. They are now trying to implement GFR estimation. They emphasize using GFR as an education tool and putting patients in a treatment category based on GFR. He cites an example of a patient whose decreased kidney function would not have been noticed were it not for the GFR estimate. In terms of implementing GFR into practice Kaiser chose to apply the term "normal" to values of 90 or over. They report the GFR as an interpretation, not as a result. There is a debate about whether or not to give age-adjusted norms. They use a formula of two standard deviations below the mean or GFR + 1/2 age less than 85. The laboratories are advised to report for African Americans and non-African Americans since it is hard to correctly identify race. A note on the report explains that this estimate assumes a steady state and is not applicable to rapidly changing function. They found that the laboratory could not do an exponential function and was defaulting to the non-Black for race. The note on the report also reads that GFR 60-89 may be normal for age > 70 although there is no real data on this yet. The report explains normal, moderately reduced levels and kidney failure. They developed a table for every two years that shows ranges that drive numbers. It shows GFR ranges and age and helps physicians classify patients by creatinine and GFR.

Kaiser also does inter-laboratory comparison testing for creatinine. They pooled a serum sample and sent it out to all the laboratories. The Cleveland Clinic did a comparison test for Kaiser Southern California and the results were good.
Kaiser's staging algorithm follows the K/DOQI Guidelines in defining CKD as kidney damage for greater than or equal to 3 months as defined by structural or functional abnormalities of the kidney, with or without decreased GFR, manifest by either pathological abnormalities or markers of kidney damage; or a GFR of less than 60 for greater than or equal to 3 months with or without kidney damage. Kaiser extends this definition to require that changes in stages also persist for more than 3 months. Their algorithm compares the current GFR to a GFR done 3 months previously and both must be out of range in the same direction. They then categorize patients based on these results. They further subdivide stage 5 into 5 transplant substages and two substages of

dialysis (based on the type of dialysis); this can be done based on the GFR. If Kaiser uses microalbuminuria as a measure more patients are classified as having CKD. Using macroscopic proteinuria yields a smaller number. Those in CKD stage 3 show a large increase in number. Kaiser has an aggressive pre-ESRD program for about 14, 000 patients. That number may increase to between 60-80 thousand. Kaiser numbers compare with those found in NHANES III so their assessment is that they are seeing those who need to be seen without getting serum creatinine levels on every patient. Currently men with creatinine values of 2 and women with creatinine values of 1.5 are referred to a nephrologist. The majority of patients seen by the nephrologists have creatinines of 4 and 5 but there is a small group of patients with creatinines of 3-4. There is a concern that they may be missing the following three groups: the elderly who have a lower risk and younger patients with diabetes and those with uncontrolled proteinuria.

There was a discussion about the large number of patients in stage 3 and the resources required to follow and manage those patients. The Kaiser data agree with NHANES III so it is reasonable to assume that that large number is real. If Kaiser has difficulty managing these patients how will other organizations find those who are normal and those who are not? How can aged patients be further classified and treated? In the Kaiser system the primary care physicians have no disincentive in referring patients to specialists; they do not lose them once they are referred, as is the case in other organizations. The Kaiser nephrologists want to see these patients. In practice they can see those who are at the highest risk even if they are at an earlier stage. For those at less risk the primary care physicians can be educated on proper treatment in a manner similar to the heart program. It is positive that these patients are being seen and having their creatinine measured. Tailoring treatment may not be too difficult: anemia can be monitored and blood pressure controlled. In the Kaiser system patients are put on an ace inhibitor (ACEi) for high blood pressure, angiotensin receptor blockers (ARB) for proteinuria or a statin if the lipids are not being managed to less than 100. It is not yet clear if age specific cut-offs for CKD stages are appropriate. Older people are at risk for many things and not all abnormalities are treated. However, there is not yet an evidence base to govern treatment when proteinuria is not present. Kaiser is building a database on this type of patient now.

IV. PANEL DISCUSSION

Panel members included representatives from the Centers for Disease Control (CDC), the Food and Drug Administration (FDA), an instrument maker, and academicians with experience in setting laboratory standards. U.S. Laboratories generally use artificial stabilized materials to evaluate their performance against proficiency testing goals established under the Clinical Laboratory Improvement Act (CLIA). As such, it is not possible to directly compare results and test methods across laboratories and definitively establish how well one method or laboratory compares to another. Therefore, great care must be taken in reviewing laboratory proficiency testing data. There is also biological variability within an individual, which must be taken into account. Biological factors contribute to serum creatinine variation on a daily, weekly and monthly basis on the order

of 3-6% in most literature reports. The intra-individual variability observed for cholesterol is discussed in the NCEP Guidelines. For cholesterol in plasma or serum, intra-individual variability is approximately twice that for creatinine, and NCEP recommends multiple measurements over weeks to a few months prior to instituting anything other than healthy lifestyle changes. The within-day variability in creatinine may be related to ketones and fasting. This variability (i.e., fasting vs. taking a measurement just after a meal) may not be that significant in impacting the computed GFR. The discussion then centered on two questions. Should something be done to standardize creatinine measurements used to calculate an estimated GFR? Should there be an effort to tighten inter-laboratory analytical variability in order to calculate estimated GFRs?

It was thought that the most straightforward approach would be to develop reference material validated by NIST, validate commutability, and make it available to manufacturers. A participant pointed out that the analytical specificity of the Jaffe reaction may be a major limiting factor in the success of standardization. Other methodological principles are available for measurement of creatinine, but they are more expensive. It is possible to standardize the calibrator to a true value of creatinine. It may be useful to work on the improving the Jaffe methods and/or to promote use of available alternative methodology (e.g., enzymatic methods), while simultaneously working on the calibration issues.

Another participant emphasized that assessing serum creatinine should be looked upon as a screening measure. If sensitivity is gained at the expense of specificity, the patient pool is narrowed at the follow-up testing. If the current method is viewed simply as a screen, it is the accuracy of the follow-up confirmatory test that is more important. The current methods for creatinine are optimized to have linearity to 20 to 30 mg/dL. They could be optimized better to produce accurate and precise values between 0 and 3 mg/dL and to report the measurement to the second decimal place. The best precision for a creatinine method in the normal range appears to be about 3 to 5%. Assay precision therefore may be inadequate for many methods. As a general rule, acceptable assay precision should have a CV of less than one-half to one-quarter of the intra-individual biological variability, which is 3 to 6% in most studies. In terms of specificity, there are many things that interfere with the Jaffe reaction, which could possibly be improved with minor increase in analytical costs. Jaffe could be an interim method until better methods with reasonable costs are developed. Some Jaffe methods compare better than others with the reference method, for example by measuring color development kinetically.

At present, NKDEP does not advocate screening for all adults. There is data to support screening of diabetics. A participant pointed out that it is more valuable to know a creatinine result is accurate in the range of 1 to 3 mg/dL than in the range > 5mg/dL from a clinical standpoint. Making the assay more accurate and precise in that lower range would have greater clinical consequences. However, this change may require reconfiguring the reagents and instruments and will take time. Presently, it should be possible to report creatinine mg/dL results to two decimal places, since all manufacturers report to two digits when they use SI units (i.e., umol/L). The time it would take to

achieve better analytical precision in the normal reference range, would probably be several years. At least one representative from an instrument company advocated using an enzymatic method calibrated to match HPLC, but enzymatic methods have not been widely accepted by laboratories due to their higher cost. The creatinine values reported by enzymatic methods are typically 0.2 mg/dL lower than those by instrument using the Jaffe method. There are different ways to try to achieve an improvement in methods in the range of 1 to 3 mg/dL. The fundamental calibration difference is important. Most assay systems use lyophilized calibration materials that has to be reconstituted. Liquid calibrators are easier to use, but are more expensive to ship and store in the laboratory (e.g., frozen or refrigerated). There are a variety of means to achieve improved accuracy and precision, but it is necessary first to establish widely acceptable bias and precision and then allow the manufacturers to develop their own plans to reach those standards. It typically takes several years for manufacturers to improve an assay system.

Creatinine may be viewed as both a screening and a monitoring test. Cystatin C is a potential test for estimation of GFR, but is very unlikely to be widely used for another couple of years. It is more difficult to do and more costly. Although, the cost is very high the Mayo Clinic does actual GFR measurement with a one-hour iothalamate clearance, but the cost is very high. Improving the serum creatinine test appears to be the best and most practical option for the present. Creatinine clearance and continued collection of urine for creatinine appears to not be the best way to proceed. It is necessary to develop a means to get true serum creatinine values. The clinical practice guidelines focus on GFR and staging by GFR. It is important to improve what is widely used now. As the creatinine assays improve, thinking will evolve and physicians will use a computed GFR as a screening tool for renal disease.

Another participant pointed out that the most important issue is educating practitioners. NKF is working on guidelines so there is no need to duplicate that effort. NKDEP advocates an annual serum creatinine for diabetics, those with hypertension, and relatives of those with ESRD using serum creatinine and computed GFR at three-year intervals. However, more clinical data is needed in support of this recommendation, especially for non-diabetics. Education of primary care physicians, NKDEP agrees, is paramount. Patients are the primary target group. They will go to their doctors and ask to be tested for kidney disease. In terms of nephrologists, it is important to reach their professional societies such as the American Society of Nephrology and the American Renal Association who in turn will educate their members. Nephrologists must learn to deal with patients early in the disease process.

FDA panel members were asked if focusing on GFR rather than on creatinine clearance constitutes a different “IVD device” from a regulatory standpoint. Is an estimated GFR a diagnostic device? The response was that the more complicated the mathematics used to make the GFR estimate, the greater chance it may be viewed by the Agency as a device. When new equations come about and are implemented clinically, the greater the clinical acceptance and use of the equations the more likely they will be viewed as a *fait accompli* when reviewed by FDA. However, use of any GFR-estimating equation will have to be addressed when it is presented to FDA.

Those who produce the creatinine results, laboratory personnel, must also be educated. When asked what their creatinine method's bias is, most do not know. Bias is a major contributor to inter-laboratory variability of serum creatinine results and hence variability of computed GFR. It is important that laboratories be educated on what they can do to monitor bias and reduce it. Laboratory personnel are focused on proficiency testing and passing that test; perhaps accrediting bodies should change requirements to a more stringent bias acceptability. It is difficult to improve the quality of test results without more stringent acceptability standards. The Clinical Laboratory Improvement Amendments of 1988 (CLIA '88) standards have very broad windows of acceptability, and the pass rate for laboratories is very high (>95%). It is difficult for proficiency test providers (e.g., CAP, AAB, etc.) to tighten laboratory standards to the point that a substantial fraction of laboratories fail, because typically when a laboratory fails it simply goes to another proficiency testing provider.

A participant pointed out that changing a proficiency testing program's standards window of acceptability for any regulated analyte such as serum creatinine, literally requires congressional action, and thus will be very difficult to change in the short term. The point is to focus laboratories on doing better by providing an incentive. The current mindset is usually that passing the proficiency test means the laboratory is doing just fine. However, laboratories can only do as well as the instruments provided by manufacturers. In Europe, there is an IVD Directive to improve the situation. In the U.S., it may be time for a new proficiency testing program to offer highly commutable materials. Industry will also need support with a means of assessing their internal calibration procedures. Currently, there is no third-party to monitor IVD manufacturing companies who may turn out calibration that is inconsistent over time. A new quality assessment program should be focused on standardization and getting the right answer to meet the NKDEP guidelines, not merely on getting the clinical laboratory accredited.

The international normalization ratio (INR) which incorporates a mean of prothrombin time results from normal healthy individuals has been quite successful in reducing inter-laboratory variability of coagulation test results used to monitor sodium warfarin administration. In October 2003, a new CAP fresh frozen serum creatinine survey will give NKDEP an opportunity to compare the MDRD-computed GFR across laboratories and instrument platforms. Data from this CAP project can help assess a creatinine version of the INR until true creatinine values are routinely reported by clinical laboratories. Results of this CAP project can also tell manufacturers how their method compares to the method used to establish the MDRD equation. Specifications also need to be developed for acceptable total error clinically, and then a laboratory standardization panel must set short- and long-term strategies. This is the pattern followed by the National Cholesterol Education program (NCEP).

There are no laboratory instruments which currently calculate an estimated GFR. This is an inefficient approach since the instruments that measure creatinine typically have no patient age, sex, and race information. A better approach is to use the laboratory or hospital information system to calculate the GFR because these systems are where the

patient information resides. There are practical issues and a panel needs to outline what those issues are and how to solve them. The laboratory standardization panel should have representation from laboratory and/or hospital IT providers, as well as laboratory instrument manufacturers. A multidisciplinary systems approach is needed before it will be possible to give clinicians the GFR they need.

In terms of method-specific bias, it is important to look at each method individually. The overall method bias may be significant, but within a specific laboratory using a given manufacturer's specific method, there may be calibration biases greater than or less than the overall method's average bias. Another participant from a healthcare system said their medical centers and clinics use instruments from several different manufacturers and the calibrations vary by specific instrument even within the same manufacturer. Giving a single bias-correction factor may be inadequate. In looking at multiple platforms, depending on the specific instrument used it is estimated they could still be off by about a factor of 10%. Thus, problems may still exist in having a patient go from one hospital or clinic to another within the same healthcare system. It is important to make the recommendations that even smaller laboratories can implement. Unless all creatinine methods used are addressed, there will still be variable bias across their laboratories.

It may be relatively easy to report creatinine values to two digits (i.e., x.xx mg/dL) in the near future. However, doing that will be market driven. Doing logarithmic transformations is a problem for some older LIS equipment. Perhaps look-up tables could be used, even possibly having a table driven method of reporting age. However, reporting race may require significant additional effort at the phlebotomy sites. If race data is missing, a participant from a reference laboratory indicated that more customer service follow-up would definitely be needed. There needs to be more market demand for reference laboratories to report GFR because undoubtedly some additional expense will be required.

A short short-term goal of a total error of the serum creatinine-based computed GFR being within plus or minus 15% of the patient's true GFR was proposed. A question related to this goal is how to partition analytical bias and analytical precision specifications. The Cholesterol Laboratory Standardization Panel looked at a bias and precision error budget. It became important to minimize precision in order to focus greater efforts on the bias budget, the most difficult to decrease. Manufacturers can work independently on improving method precision in the normal and near-normal range for serum creatinine. It is better to give manufactures more flexibility in the error budget to work on bias. Simulations can help determine the impact of inter-individual variability and whether multiple blood samples spread across weeks or months are needed. If there is an average of 3-6% intra-individual variability, it may be necessary to measure two samples from an individual at different points in time. Theoretically, a goal of 15% for total error on serum creatinine encompasses bias and imprecision for that individual measurement. A goal of 15% may be too strict, and another participant suggested that twice that or 30% would be more realistic. More work is required on what the analytical precision and bias goals should be.

More input from the diagnostic manufacturers has been helpful. How can NKDEP reach this group and those who decide error budgets? With the lipid standardization program and the HbA1c program appropriate clinically oriented bodies spoke out about what was needed. In the case of the lipids, a NCEP formal Laboratory Standardization Panel was formed which gave specifications and recommendations on what needed to be done and how proceed in accomplishing specific goals. For cholesterol, there were recommendations on what government, manufacturers, and laboratories each needed to do. A similar laboratory standardization panel for creatinine could develop more recommendations that manufacturers can take to their internal corporate QC groups. Such a panel might also develop laboratory informatics recommendations, rather than having a disjointed effort.

The goals for lipids were set with a couple of timeframes: an interim three-year goal, then longer term seven-year goal with the idea of improving incrementally in stages. These were goals that allowed adequate time for all manufacturers to respond. The components of imprecision that should be reviewed include lot-to-lot stability for reagents as well as for total performance. Short-term within batch also makes a difference. With respect to short-term or day-to-day variation, many methods may be good enough. However, it is unknown whether the same estimates and precision can be attained when using commutable material. An estimate of how many replicates are needed in order to deal with biological variation is also an outstanding item. The performance of the Jaffe method is compromised by analytical non-specificity of the method. The biological variation is also affected when an individual fasts, takes medication, has a disease, etc. A study comparing the intra-individual variation observed with the enzymatic method versus the Jaffe method might be important to explain some of this variation, and possibly offer an approach to control or minimize it. The specificity of the enzymatic method is an improvement. While it is possible to improve the analytical methods, the inherent biological variation probably cannot be improved, but it does need to be understood. In terms of total error, if one is dealing with a GFR of 50, for example, it would be reassuring to know that a reported GFR is plus or minus 10 at most of the patient's true mean GFR. This represents a total error goal of plus or minus 20%. This might call for two serum creatinine measures. One recommendation is that analytical imprecision should be less than one-fourth of the biological variation, which for creatinine is an analytical imprecision of only 1%. That improvement might be achieved in two years, once an improvement program is initiated. The cholesterol program is still working toward the goals that it set over ten years ago. Precision can likely be improved in the short-term for creatinine in the normal range.

There are 5 work group tasks as subjects of further discussion:

1) Define an acceptable highest-level reference measurement procedure for serum creatinine.
2) Define one or more mid-level reference measurement procedures for serum creatinine.
3) Define a list of acceptable reference measurement laboratories that can perform #1 and/or #2 above.

4) Agree upon a commutable reference material that can be used to establish traceability to field measurement procedures from mid- and highest-level reference measurement procedures.
5) Establish a program for periodic evaluation of the ability of clinical laboratories and specific laboratory methods to report true creatinine values.

The first three points above concern reference measurement procedures and reference materials. A Joint Committee on Traceability in Laboratory Medicine (JCTLM) exists which could take on this task. This is an international group stemming from ISO Technical Committee 212 and European IVD Directive requirements to define reference methods and materials so that manufacturers can establish traceability of their field measurement procedures to high-level reference measurement procedures. A statement on commutability will also come from this committee. JCTLM has a Working Group 1: Reference Materials and Reference Procedures. NKDEP could interact with this Working Group on tasks 1, 2 and 3 above. It is important to work with an international group to prevent developing multiple national or regional reference methods that potentially differs from an international one. NIST is involved in the international activity. The JCTLM working group needs input on setting priorities based on clinical need. It was suggested that NKDEP and a laboratory standardization panel be set up (or a steering committee) that could draft a statement to help prioritize the creatinine efforts of the JCTLM in international standards related to creatinine. The Chairman of the JCTLM is Professor Joseph H H Thijssen (email: j.thijssen@lab.azu.nl). The Working Group 1 has Joint Chairs, one of whom is Dr. Willie E. May of NIST (willie.may@nist.gov).

Funding to set standards and develop a commutable reference material was discussed. It is unclear who should make the material. There are European laboratories that could provide the material, but cost is an issue. CAP no longer provides calibration materials because it was not cost effective. It was agreed that some group, potentially a NKDEP laboratory standardization panel, has to evaluate needs, costs, and potential sources of funding.

A program to evaluate the ability of clinical laboratories periodically and their methods' ability to report true creatinine values should be established. However, it was suggested that this might be done as part of JCTLM. However, a long-term QA maintenance function is necessary to sustain the effectiveness of the laboratory and manufacturer's standardization programs. If this is a serious public heath problem, then it needs to be maintained and quality-assured. This may be a government function.

In terms of evaluating the GFR equations, it was suggested that it would be wise to wait for JCTLM to decide the higher order method(s) for creatinine. If a standard is set now and later differs from the JCTLM, the program would have to be redone. Confusion should be kept to a minimum.

V. CONCLUSION

Dr. Hostetter thanked participants. This meeting was intended to develop the next steps. The first meeting of the education program was 6 months ago. As this process moves ahead there will be improvements in estimates of GFR. Market forces drive this process. Market representatives participated today and the market is changing. The number of patients with ESRD has doubled. NKDEP is told by Kaiser, Coventry and disease management organizations that there is interest in finding those with kidney disease earlier in the disease process.

Dr. Hostetter asked attendees to send him their ideas on how the National Institutes of Health, the National Institute of Diabetes and Digestive and Kidney Diseases and the National Kidney Disease Education Program can move this effort forward. Educational materials available for download can be developed. As a next step a panel will develop goals and actions to achieve those goals. This is a process, which is necessary and wanted by organized medicine.

VI. ATTACHMENTS

1. Agenda
2. Participant List